

Tilburg University

## Bayesian marginal equivalence of elliptical regression models

Osiewalski, J.; Steel, M.F.J.

*Publication date:*  
1991

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Osiewalski, J., & Steel, M. F. J. (1991). *Bayesian marginal equivalence of elliptical regression models*. (CentER Discussion Paper; Vol. 1991-19). Unknown Publisher.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### Take down policy

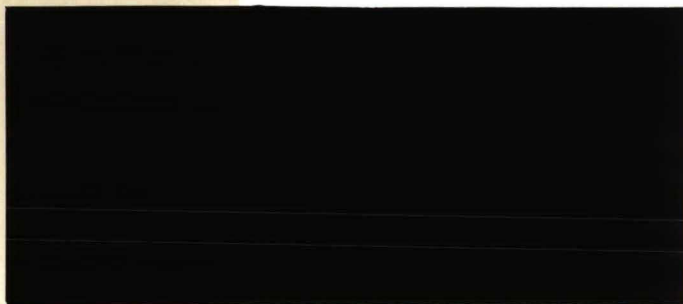
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.


CBM  
CBM  
R  
8414  
1991

**entER**  
for  
mic Research

19

# Discussion paper





No. 9119

**BAYESIAN MARGINAL EQUIVALENCE OF  
ELLIPTICAL REGRESSION MODELS**

by Jacek Osiewalski  
and Mark Steel

R20

32045

May 1991

ISSN 0924-7815

# BAYESIAN MARGINAL EQUIVALENCE OF ELLIPTICAL REGRESSION MODELS\*

Jacek Osiewalski  
Academy of Economics  
Kraków, Poland

Mark F.J. Steel\*\*  
Tilburg University  
The Netherlands

The use of proper prior densities in regression models with multivariate non-Normal elliptical error distributions is examined when the scale matrix is known up to a precision factor  $\tau$ , treated as a nuisance parameter. Marginally equivalent models preserve the convenient predictive and posterior results on the parameter of interest  $\beta$  obtained in the reference case of the Normal model and its conditionally natural conjugate gamma prior. Prior densities inducing this property are derived for two special cases of non-Normal elliptical densities representing very different patterns of tail behaviour. In a linear framework, so-called semi-conjugate prior structures are defined as leading to marginal equivalence to a Normal data density with a fully natural conjugate prior.

\* A first version of this paper was written while the first author was a visiting fellow at CORE, Louvain-la-Neuve. The second author holds a research fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW). We are indebted to Michel Mouchart and anonymous referees for very helpful comments and suggestions, though the usual disclaimer applies.

\*\* Department of Econometrics, Tilburg University,  
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

## 1. Introduction

Part of the literature in Bayesian econometrics has been directed towards broadening the distributional assumptions on the error terms of the multiple regression model. Zellner (1976) considered Student  $t$  errors and concluded that inference still remains relatively simple with diffuse priors. Jammalamadaka et al. (1987), Chib et al. (1988) and Osiewalski (1991) considered errors distributed as scale mixtures of Normals and stated that under certain improper prior assumptions both prediction and posterior inference is unaffected by such departures from Normality.

Under **improper** priors, these results were generalized to any multivariate elliptical data density in Osiewalski and Steel (1990). They showed that it suffices to single out a scalar precision factor  $\tau$  on which we specify a Jeffreys' type prior to obtain full robustness within the entire family of multivariate elliptical sampling models. This robustness property holds for both predictive and for posterior results on the parameters other than  $\tau$ . If we, however, insist on using **proper** prior structures, the results of Zellner (1976) already suggest that such robustness no longer occurs.

We consider two parametric families of sampling densities  $P$  and  $P_*$  with the parameter of interest  $\beta$  in common and different sets of nuisance parameters. Bayesian models from both families are called **marginally equivalent** if prior densities are such that they lead to the same posterior inference on  $\beta$  and the same predictive inference. In particular, we take  $P$  to be the class of multivariate **elliptical** data densities with location parameter  $\beta$  and nuisance parameter  $\delta$  involving  $\tau$ . For  $P_*$  we choose the usual Normal sampling model with the same location vector and with nuisance precision factor  $\rho$ . A convenient reference prior  $p_*(\rho|\beta)$  is the conditional natural conjugate gamma density, and we examine which (proper) priors on  $\delta$ , given  $\beta$ , make a non-Normal member of  $P$  marginally equivalent to the Normal model with gamma prior. For two leading examples, the Student  $t$  case and the model which is **uniform over an ellipsoid**, such conditional priors of  $\delta$  are derived in closed form. These special cases correspond to, respectively, thicker tails than the Normal, and truncated

tails, and thus require very different priors to offset this tail behaviour. The effect of the heavy tails in the Student data density is neutralized by prior restrictions on the parameter space. In the opposite case of uniformity over the interior of an ellipsoid, the absence of sampling tails is compensated by a conditional prior density of  $\delta$  which has a thicker tail than the reference gamma density.

In the linear regression model, an even more convenient prior structure is the particular Student-gamma form which is natural conjugate for both  $\beta$  and  $\rho$  under Normality. For alternative linear elliptical sampling processes, any prior on  $(\beta, \delta)$  that induces marginal equivalence with this Normal-natural conjugate model will be called **semi-conjugate**. Mimicking the behaviour of this most popular reference model is seen to imply some potentially severe restrictions.

For convenience, probability density functions not explicated in the course of the paper, are grouped in an appendix.

## 2. Marginally equivalent Bayesian models

Consider a parametric family  $P = \{p(y|\beta, \delta) : \beta \in B, \delta \in \Delta\}$  of probability densities for a vector observation  $y$ , where  $\beta$  is a parameter of primary interest and  $\delta$  is a nuisance parameter. Suppose also that there is another parametric family  $P_* = \{p_*(y|\beta, \varphi) : \beta \in B, \varphi \in \Phi\}$  for  $y$  in which  $\beta$  plays the same role as in  $P$  (e.g.,  $\beta$  is a location parameter for both  $P$  and  $P_*$ ), but  $\varphi$  need not be linked to  $\delta$  (even the dimensions can differ). Now consider the two Bayesian models, i.e. joint densities for observations and parameters,

$$p(y, \beta, \delta) = p(y|\beta, \delta)p(\beta, \delta), \quad (2.1)$$

$$p_*(y, \beta, \varphi) = p_*(y|\beta, \varphi)p_*(\beta, \varphi), \quad (2.2)$$

where  $p(\beta, \delta)$  and  $p_*(\beta, \varphi)$  are prior densities for the parameters of  $P$  and  $P_*$ , respectively.



If  $\beta$  has the same interpretation in both parametric families, we can regard it as having a reality independent of the choice of  $P$  or  $P_*$ , and thus we will naturally require that the marginal prior density of  $\beta$  does not depend on the choice of the sampling family, i.e.

$$p(\beta, \delta) = p(\beta)p(\delta|\beta) \quad (2.3)$$

and

$$p_*(\beta, \varphi) = p(\beta)p_*(\varphi|\beta) \quad (2.4)$$

Suppose that  $p_*(y, \beta, \varphi)$  has a particularly convenient form. The issue is then whether we can use this convenient model for Bayesian marginal posterior inference about  $\beta$  and for predictive inference when  $y$  comes from a density in  $P$  rather than a density in  $P_*$ . For this to be valid, the Bayesian models in (2.1) and (2.2) must be **marginally equivalent** for  $y$  and  $\beta$ , i.e.

$$p(y, \beta) = p_*(y, \beta), \quad (2.5)$$

where  $p(y, \beta) = \int_{\Delta} p(y|\beta, \delta)p(\beta, \delta)d\delta$ ,

$$p_*(y, \beta) = \int_{\Phi} p_*(y|\beta, \varphi)p_*(\beta, \varphi)d\varphi.$$

Under (2.3) and (2.4), (2.5) reduces to the requirement that the marginalized likelihoods

$$p(y|\beta) = \int_{\Delta} p(y|\beta, \delta)p(\delta|\beta)d\delta \quad (2.6)$$

and

$$p_*(y|\beta) = \int_{\Phi} p_*(y|\beta, \varphi)p_*(\varphi|\beta)d\varphi \quad (2.7)$$

be identical, i.e.

$$p(y|\beta) = p_*(y|\beta). \quad (2.8)$$

In this paper we assume that  $P$  is the class of  $n$ -variate non-Normal elliptical data densities with location vector  $h(X, \beta)$  and scale  $\tau^{-1}V$ , redefining  $\delta$  as  $\delta = (\tau, \nu)$

$$p(y|\beta, \delta) = f_g^n(y|h(X, \beta), \tau^{-1}V) = |V|^{-\frac{1}{2}} \tau^{\frac{n}{2}} g\{[y-h(X, \beta)]' \tau V^{-1} [y-h(X, \beta)]\}, \quad (2.9)$$

where  $g(\cdot)$  is a known nonnegative function indexed by  $\nu$  such that  $u^{\frac{n}{2}-1} g(u)$  is integrable in  $u$  over  $R_+$ . The latter requirement is shown in e.g. Kelker (1970) and Dickey and Chen (1985) to be necessary and sufficient for properness of (2.9).  $g(\cdot)$  essentially controls tail behaviour. Note that non-Normality of (2.9) means that  $g(u)$  is not exponential in  $u \in R_+$ . We also assume throughout the paper that the alternative family  $P_*$  consists of the Normal densities

$$\begin{aligned} P_*(y|\beta, \varphi) &= f_N^n(y|h(X, \beta), \varphi^{-1}V) = \\ &= (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \varphi^{\frac{n}{2}} \exp\{-\frac{\varphi}{2} [y-h(X, \beta)]' V^{-1} [y-h(X, \beta)]\}. \end{aligned} \quad (2.10)$$

Thus, in the case considered in the paper,  $\tau$  and  $\varphi$  are positive scalar parameters (precision parameters) and  $\nu \in N$  which may be empty. In both (2.9) and (2.10),  $V$  is a known  $n \times n$  PDS matrix, and  $h$  is a known vector function of the matrix  $X$  and of  $\beta \in B \subseteq R^k$ . The regression models condition on  $X$  which is independent of all the parameters in the implied conditional models. A convenient concept that ensures the latter condition is a Bayesian cut [see Florens and Mouchart (1985)]. As conditioning on  $X$  will be maintained throughout the analysis, it will not be explicit in the notation. Remark that, in (2.9) and (2.10), the location of ellipsoids is entirely determined by  $\beta$  (given  $X$ ), which has an unambiguous interpretation, irrespective of the parametric family we choose.

Definition 1: any elliptical sampling model from  $P$  in (2.9) together with a prior on the nuisance parameter  $p_g(\delta|\beta)$  is **marginally equivalent** to a Normal model from  $P_*$  in (2.10) with the prior  $p_*(\varphi|\beta)$  if, under (2.3) and



(2.4), the marginalized likelihoods  $p(y|\beta)$  and  $p_*(y|\beta)$  in (2.6) and (2.7) coincide.  $\square$

An important example of marginal equivalence of Bayesian models is given by Osiewalski and Steel (1990) who show that the data density (2.9) and the improper prior structure

$$p(\beta, \delta) = p(\tau)p(\beta, \nu) = c\tau^{-1}p(\beta, \nu), \quad \tau \in \mathbb{R}_+, \beta \in B, \nu \in N, \quad (2.11)$$

where  $c$  is any positive constant and  $p(\beta, \nu)$  is integrable in  $\nu$  over  $N$ , lead to the marginalized likelihood

$$p(y|\beta) = c\Gamma\left(\frac{n}{2}\right)|V|^{-\frac{1}{2}}\{[y-h(X, \beta)]'V^{-1}[y-h(X, \beta)]\}^{-\frac{n}{2}} \quad (2.12)$$

for any  $g(\cdot)$  not indexed by  $\tau$ . Note that (2.12) is the same as the marginalized likelihood obtained from (2.10) under the prior structure

$$p_*(\beta, \varphi) = p_*(\varphi)p(\beta) = c\varphi^{-1}p(\beta), \quad \varphi \in \mathbb{R}_+, \beta \in B. \quad (2.13)$$

Since for the Bayesian models (2.9), (2.11) and (2.10), (2.13) marginal equivalence holds for any such  $g(\cdot)$  in  $P$ , Osiewalski and Steel (1990) arrive at **robustness** of posterior and predictive results with respect to departures from Normality within a broad class of multivariate elliptical data densities.

In this paper we are also looking for Bayesian marginal equivalence of (2.9) and (2.10), but under **proper** prior densities  $p_g(\delta|\beta)$ . As could be expected, the results will be much more modest than under the improper priors in (2.11) and (2.13).

### 3. Marginal equivalence under proper priors on $\delta$ given $\beta$

Consider the two alternative sampling families  $P$  and  $P_*$  in (2.9) and (2.10). Given  $\beta$ , the precision parameter  $\varphi$  of the Normal data density (2.10) is now assigned the very convenient natural conjugate gamma prior

$$p_*(\varphi|\beta) = f_G(\varphi|\frac{a}{2}, \frac{d_\beta}{2}), \quad (3.1)$$

where  $a$  is a positive constant, and  $d_\beta$  is a known positive function of  $\beta$  ( $d_\beta$  may be a constant function  $d$ ). Our interest is in a conditional prior  $p_g(\delta|\beta)$  of  $\delta$  in (2.9) such that the Bayesian models

$$p(y, \beta, \delta) = f_g^n(y|h(X, \beta), \tau^{-1}V)p_g(\delta|\beta)p(\beta) \quad (3.2)$$

and

$$p_*(y, \beta, \varphi) = f_N^n(y|h(X, \beta), \varphi^{-1}V)f_G(\varphi|\frac{a}{2}, \frac{d_\beta}{2})p(\beta) \quad (3.3)$$

are marginally equivalent. If such a prior  $p_g(\delta|\beta)$  exists, it must fulfil (2.8) which now specializes to the equation

$$\int_{\Delta} f_g^n(y|h(X, \beta), \tau^{-1}V)p_g(\delta|\beta)d\delta = f_S^n(y|a, h(X, \beta), \frac{a}{d_\beta}V^{-1}), \quad (3.4)$$

where the RHS is an  $n$ -variate Student  $t$  as defined in the Appendix. For non-Normal elliptical densities, this  $p_g(\delta|\beta)$  will generally differ from  $p_*(\varphi|\beta)$  in (3.1), and will certainly not always exist in closed form from (3.4).

Since both  $\varphi$  and  $\delta$  are nuisance parameters, the most straightforward way to elicit values for the hyperparameters  $a$  and those in  $d_\beta$  in (3.1) is through the implied properties of the error vector  $\epsilon = y - h(X, \beta)$ , which, given  $X$  and  $\beta$ , follows a Student  $t$  density with  $a$  degrees of freedom, mean zero if  $a > 1$  and  $\text{Var}(\epsilon|X, \beta) = d_\beta/(a-2)V$  if  $a > 2$ . Prior conditional moments of  $\epsilon$  exist up to  $a$ , which is the prior counterpart of  $n$  and conveys the strength of prior beliefs. Values for the hyperparameters in the prespecified form of  $d_\beta$  (or values for  $d$ ) can then be elicited by comparing  $\text{Var}(\epsilon|X)$  to the unconditional variance of  $y$  approximated using the sampling variance [see Richard and Steel (1988, Appendix D)]. If we wish to allow for values of  $0 < a < 2$ , the elicitation procedure can be based on the precision parameter  $\varphi$  directly, using  $E(\varphi|\beta) = a/d_\beta$  and  $\text{Var}(\varphi|\beta) = 2a/d_\beta^2$ . Nevertheless, since these moments will generally not

coincide with those of  $\tau$  for marginally equivalent models, we prefer to link the elicitation process to  $\epsilon$ , which has an unambiguous meaning, regardless of the particular model. Of course, if  $\delta$  were a parameter of interest, then the prior  $p_g(\delta|\beta)$  that induces marginal equivalence for a particular choice of non-Normal elliptical process, may not be judged a reasonable description of prior beliefs. However, if  $\delta$  is a nuisance parameter, the model user will typically not possess prior information on it.

In the following subsections we give two very different examples of tail behaviour where closed form solutions of  $p_g(\delta|\beta)$  can be derived. In practice, such solutions can often be suggested by considering one-to-one transformations from  $(y, \beta, \varphi, \nu, v)$  to  $(y, \beta, \tau, \nu, z)$ , where  $v$  and  $z$  are auxiliary variables that allow us to exploit the fact that the Student density in Subsection 3.1 is a scale mixture (through  $z$ ) of Normals, and, inversely, the reference Normal can be represented as a scale mixture (through  $v$ ) of densities uniform over an ellipsoid as used in Subsection 3.2.

### 3.1. The multivariate t data density

Consider the standard  $n$ -variate Student  $t$  data density with  $\nu$  degrees of freedom:

$$f_g^n(y|h(X, \beta), \tau^{-1}V) = f_S^n(y|\nu, h(X, \beta), \tau V^{-1}) \quad (3.5)$$

which is of the form (2.9) with

$$g(u) = \frac{\Gamma(\frac{n+\nu}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{n/2}} \left(1 + \frac{u}{\nu}\right)^{-\frac{n+\nu}{2}},$$

depending upon the parameter  $\nu \in \mathbb{R}_+$ . Tails are now thicker than in the Normal case (2.10) and moments exist up to (not including)  $\nu$ .

Proposition 1: the Student sampling model in (3.5) combined with the following conditional prior:

$$\begin{aligned}
p_g(\delta|\beta) &= p_g(\tau|\nu, \beta) p_g(\nu|\beta) \\
&= f_B(\tau|\frac{a}{2}, \frac{\nu-a}{2}, \frac{\nu}{d_\beta}) p_g(\nu|\beta)
\end{aligned} \tag{3.6}$$

where  $p_g(\nu|\beta)$  is a proper density which is zero for  $\nu \leq a$ , and with any (possibly improper)  $p(\beta)$ , is marginally equivalent with  $p_*(y, \beta, \varphi)$  in (3.3).

Proof: it suffices to check (3.4). In particular, we obtain

$$\begin{aligned}
p(y|\beta) &= \int_{\Delta} f_S^n(y|\nu, h(X, \beta), \tau V^{-1}) \\
&\quad f_B(\tau|\frac{a}{2}, \frac{\nu-a}{2}, \frac{\nu}{d_\beta}) p_g(\nu|\beta) d\nu d\tau.
\end{aligned} \tag{3.7}$$

After the variable transformation from  $\tau$  to  $\xi = \frac{\tau}{\nu} / (1 - \frac{\tau d_\beta}{\nu})$ , we can integrate out  $\xi$  using a beta prime density (see Appendix), and are left with

$$p(y|\beta) = f_S^n(y|a, h(X, \beta), \frac{a}{d_\beta} V^{-1}) \int_N p_g(\nu|\beta) d\nu, \tag{3.8}$$

so that  $\nu$  can only be updated through its prior links with  $\beta$  and (3.4) is seen to hold given the properness of  $p_g(\nu|\beta)$ .  $\square$

The beta prior in (3.6) restricts the parameter space of  $\delta$  to  $(0, \nu/d_\beta)$ . For marginal inference on  $(y, \beta)$  this exactly compensates the influence of the heavy tails in (3.5) and leads to the same results as in the Normal model with gamma prior  $p_*(\varphi|\beta)$ . By choosing  $a$  to be a small enough positive number, we can span the entire Student  $t$  class, even including Cauchy densities ( $\nu = 1$ ). The parameter  $\nu$ , which does not appear in the Normal model, can never be updated "directly" in the marginally equivalent model (3.5)-(3.6). See Chib et al. (1991) for a related discussion. The conditional posterior density of  $\tau$  given  $\beta$  and  $\nu$ , however, does not retain the form of the prior, but can be written as

$$p_g(\tau|\beta, \nu, y) \propto f_B(\tau|\frac{a}{2}, \frac{\nu-a}{2}, \frac{\nu}{d_\beta}) f_{IB}(\tau|\frac{\nu}{2} - 1, \frac{n}{2} + 1, \frac{\nu}{s_\beta}), \tag{3.9}$$

where

$$s_{\beta} = [y - h(X, \beta)]' V^{-1} [y - h(X, \beta)]. \quad (3.10)$$

Prior independence between  $\beta$  and  $\varphi$  in (3.3) amounts to taking  $d_{\beta} = d$ , a positive constant, and renders the beta prior in (3.6) independent of  $\beta$ . As  $a \rightarrow 0$  and  $d \rightarrow 0$ , the kernel of this beta prior becomes proportional to  $\tau^{-1}$  (for  $\tau \in \mathbb{R}_+$ ), and the Student  $t$  marginalized likelihood in (3.8) becomes proportional to (2.12). Finally, as  $\nu \rightarrow \infty$  the sampling models in (2.10) and (3.5) become indistinguishable and the prior of  $\tau$  in (3.6) indeed tends to the gamma prior of  $\varphi$  found in (3.1).

### 3.2. Uniformity over ellipsoid

Instead of the Student tails, which are thicker than the reference Normal ones, let us now consider a case with truncated tails. In particular, we choose the special case of a multivariate Pearson Type II distribution [see e.g. Johnson (1987)], defined by taking  $g(u)$  nonzero and constant for  $u \in [0, 1]$  and zero otherwise. Then (2.9) becomes

$$f_{\beta}^n(y|h(X, \beta), \tau^{-1}V) = \begin{cases} \Gamma(\frac{n}{2} + 1) (\frac{\tau}{n})^{n/2} |V|^{-\frac{1}{2}} & \text{if } \tau s_{\beta} < 1, \\ 0 & \text{otherwise,} \end{cases} \quad (3.11)$$

where  $s_{\beta}$  was defined in (3.10). The data density (3.11) spreads the probability mass evenly over the ellipsoid  $\{y \in \mathbb{R}^n : \tau s_{\beta} < 1\}$  and thus possesses no tails. In this subclass of  $P$  the parameter  $\nu$  is absent, so that  $\delta = \tau$ .

Proposition 2: the Pearson II sampling model in (3.11) combined with the following conditional prior:

$$p_{\beta}(\delta|\beta) = p_{\beta}(\tau|\beta) = f_{IB}(\tau|\frac{n}{2} + 1, \frac{a}{2}, d_{\beta}^{-1}), \text{ and any } p(\beta), \quad (3.12)$$

is marginally equivalent with  $p_{\bullet}(y, \beta, \varphi)$  in (3.3).



Proof: we now calculate

$$p_g(y|\beta) = \int_0^{s_\beta^{-1}} \Gamma\left(\frac{n}{2} + 1\right) \left(\frac{\tau}{n}\right)^{\frac{n}{2}} |v|^{-\frac{1}{2}} f_{IB}\left(\tau\left|\frac{n}{2} + 1, \frac{a}{2}, d_\beta^{-1}\right.\right) d\tau, \quad (3.13)$$

and apply the transformation  $\tau \rightarrow \zeta$  with  $\zeta = \tau d_\beta / (1 + \tau d_\beta)$ , which allows analytical integration of  $\zeta$  and results in (3.4).  $\square$

As the data density itself is now restricted, there is no need to restrict the parameter space in order to obtain the reference result  $p_*(y, \beta)$ . Indeed, the tail of (3.12) is even much thicker than for its gamma counterpart in the Normal case. Note that the inverted beta prior in (3.12) will be truncated by the sampling model to give the following posterior:

$$p_g(\tau|\beta, y) = \begin{cases} \propto f_{IB}(\tau|1, \frac{n+a}{2}, d_\beta^{-1}) & \text{if } \tau < s_\beta^{-1}, \\ 0 & \text{otherwise.} \end{cases} \quad (3.14)$$

Again, if we assume prior independence by taking  $d_\beta = d$ , the kernel of the inverted beta prior (3.12) approaches  $\delta^{-1}$  (for  $\delta \in \mathbb{R}_+$ ) as both  $a$  and  $d$  go to zero, and the marginalized likelihood in (3.13) becomes proportional to (2.12).

#### 4. Linear regression and semi-conjugate priors

Assume that the prior density in the reference Bayesian model (3.3) takes the form

$$p_*(\beta, \varphi) = f_G(\varphi|\frac{k+e}{2}, \frac{1}{2} [f + (\beta - \bar{\beta})' A (\beta - \bar{\beta})]) f_S^k(\beta|e, \bar{\beta}, \frac{e}{f} A), \quad (4.1)$$

where  $e$  and  $f$  are positive constants,  $\bar{\beta}$  is a  $k \times 1$  vector and  $A$  is a PDS  $k \times k$  matrix. Note that  $d_\beta$  defined implicitly in (4.1) is not constant in  $\beta$ , and



thus precludes prior independence, and  $a = k+e$  is greater than the dimension of  $\beta$ . Of course, (4.1) is the well-known Student-gamma prior (or Normal-gamma in the alternative factorization), natural conjugate for both  $\beta$  and  $\varphi$  in the linear case, i.e. when  $h(X, \beta) = X\beta$ .

Definition 2: any prior density  $p_g(\beta, \delta)$  which makes the Bayesian model

$$p_g(y, \beta, \delta) = f_g^n(y | X\beta, \tau^{-1}V) p_g(\beta, \delta) \quad (4.2)$$

marginally equivalent to

$$p_*(y, \beta, \varphi) = f_N^n(y | X\beta, \varphi^{-1}V) p_*(\beta, \varphi) \quad (4.3)$$

where  $p_*(\beta, \varphi)$  is as in (4.1), will be called **semi-conjugate**.  $\square$

Semi-conjugate priors exactly preserve the simple Student t forms of the marginal prior, posterior and predictive densities of  $\beta$  and  $y$  which are obtained in (4.3). Semi-conjugate priors for the types of data distributions considered in Subsections 3.1 and 3.2 can immediately be obtained by taking in (3.2), (3.6) and (3.12) the same  $a$ ,  $d_\beta$  and  $p(\beta)$  as in (4.1). Only the Student t density case will be discussed in some detail. From (3.6) and (4.1) the prior

$$p_g(\beta, \delta) = f_S^k(\beta | e, \bar{\beta}, \frac{e}{f}, A) f_B(\tau | \frac{k+e}{2}, \frac{\nu-k-e}{2}, \frac{\nu}{f + (\beta - \bar{\beta})' A (\beta - \bar{\beta})}) p_g(\nu | \beta) \quad (4.4)$$

with  $p_g(\nu | \beta)$  proper and zero for  $\nu \leq k+e$ , is semi-conjugate for the Student t data density

$$f_g^n(y | X\beta, \tau^{-1}V) = f_S^n(y | \nu, X\beta, \tau V^{-1}) \quad (4.5)$$

The form of the semi-conjugate prior (4.4) shows that severe restrictions must be put on the Bayesian model (4.2) with the data density (4.5) if we want to mimic (for marginal inference on  $\beta$  and  $y$ ) the convenient reference model (4.3). Firstly, the implied marginal density of  $\tau$  is nonzero over  $(0, \nu/f)$  and the conditional densities of  $\beta$ , given  $\nu$  and values of

$\tau \in (0, \nu/f)$ , are nonzero over ellipsoids  $(\beta - \bar{\beta})' A (\beta - \bar{\beta}) < \frac{\nu}{\tau} - f$ . Thus, an upper bound  $\nu/f$  is put on the precision parameter  $\tau$ , and  $\beta$  values far from the prior mean (in the metric induced by  $A$ ) are allowed only for very small values of  $\tau$ , i.e. for noisy data processes, or for large values of  $\nu$ , i.e. densities close to Normality. Secondly, there is a lower bound on the degrees of freedom of the Student  $t$  sampling process in (4.6), namely  $\nu > k + e$ . Very thick (e.g. Cauchy) tails are ruled out. If the tails become too thick, i.e. if  $\nu \leq k$ , even restricting the parameter space of  $(\beta, \delta)$  no longer suffices to obtain the same results for  $p(y, \beta)$  as in (4.3). Even for very thick tails of (4.5), however, it is possible to obtain marginal equivalence (see Subsection 3.1) and thus mimic the Bayesian results for  $y$  and  $\beta$ , but then outside the natural conjugate framework (4.1). For example, by Proposition 1 the Normal data density under the Student-gamma prior,

$$p_*(y, \beta, \varphi) = f_N^n(y | X\beta, \varphi^{-1}V) f_S^k(\beta | e, \bar{\beta}, A) f_G(\varphi | \frac{a}{2}, \frac{d_\beta}{2}), \quad (4.6)$$

is marginally equivalent to the Student  $t$  data density with Student-beta prior

$$p(y, \beta, \delta) = f_S^n(y | \nu, X\beta, \tau V^{-1}) f_S^k(\beta | e, \bar{\beta}, A) f_B(\tau | \frac{a}{2}, \frac{\nu-a}{2}, \frac{\nu}{d_\beta}) p_g(\nu | \beta) \quad (4.7)$$

where  $p_g(\nu | \beta)$  is proper and only nonzero for  $\nu > a > 0$ . Here we have one more free hyperparameter than in the semi-conjugate prior (4.4), and, therefore,  $\nu$  need not be related to  $k$ . If we also take  $d_\beta = d$  and  $p_g(\nu | \beta) = p_g(\nu)$  we have prior independence between  $(\tau, \nu)$  and  $\beta$ , and the only remaining restriction on the parameter space will be that  $\tau \in (0, \nu/d)$ , which will become less binding if  $p_g(\nu)$  puts more mass on large values of  $\nu$ .

## 5. Conclusion

The use of Bayesian regression analysis in practice often relies on the Normal sampling model and its natural conjugate prior structure, since this leads to predictive and posterior densities with convenient properties. We ask whether the aspects which are typically of interest carry over to the general class of elliptical regression models. In particular, we examine the marginal equivalence for  $(y, \beta)$  of non-Normal elliptical sampling models to the Normal model with a convenient gamma prior on the precision factor  $\varphi$ , which is natural conjugate given  $\beta$ . For linear models, the specific prior structure that ensures marginal equivalence under a fully natural-conjugate density for  $(\beta, \varphi)$  in the Normal model is called semi-conjugate. The latter is of particular interest since it completely preserves the very convenient predictive and posterior results for  $\beta$ .

As already mentioned in Section 2, much stronger robustness results can be achieved if one allows for improper Jeffreys' type prior densities on the nuisance precision parameter  $\tau$  of the elliptical model (see Osiewalski and Steel 1990). However, the analysis is similar in the sense that the difference between members of the elliptical class is entirely isolated in  $\tau$ . Under a Jeffreys' prior on  $\tau$  only the conditional posterior on  $\tau$  is affected by the choice of elliptical sampling model and the inference on  $(y, \beta)$  is the same whatever the model chosen. The price to pay for restricting attention to proper prior families of the nuisance parameter  $\delta = (\tau, \nu)$  is that the robustness results are more modest. Not just the posterior, but also the prior of  $\delta$  will now vary over elliptical models. A specific prior linked to a particular non-Normal elliptical model will exactly mimic the marginal results for  $(y, \beta)$  that the natural conjugate prior structure produces with the Normal model. Differences in tail behaviour of the sampling model are entirely compensated by the properties of the conditional prior of  $\delta$ . The Bayesian user can then clearly isolate the consequences of deviating from Normality within the elliptical class of sampling models. The effect of heavy tails in the Student t case, for example, is neutralized by the beta form of  $p_g(\tau|\beta)$ , which restricts the parameter space. In the case of truncated

tails with uniformity over the interior of an ellipsoid, marginal equivalence requires an inverted beta density with a thicker tail than the natural conjugate gamma prior. So both prior and posterior distributions of  $\delta$  given  $\beta$  vary here with the choice of sampling model, allowing the rest of the analysis to remain unaffected.

If a marginally equivalent model is acceptable to a practitioner, she can simply base her inference concerning  $\beta$  and  $y$  on the standard formulae and computer programs for the Normal model, even though she has assumed a non-Normal elliptical data density. Especially semi-conjugacy can imply potentially severe restrictions, however, and a particular member of the non-Normal elliptical class always needs to be selected. For these reasons, we should advise the practitioner who wishes to deviate from Normality to first consider using a Jeffreys' diffuse prior on the precision parameter, before contemplating the use of a proper prior density. Of course, if she has relatively strong prior information on the properties of the error term, this should be included in the analysis and then marginal equivalence may indeed prove very useful.

# Appendix. Probability density functions

A  $k$ -variate Normal density on  $x \in \mathbb{R}^k$  with mean vector  $b \in \mathbb{R}^k$  and PDS  $k \times k$  covariance matrix  $C$ :

$$f_N^k(x|b,C) = [(2\pi)^k |C|]^{-\frac{1}{2}} \exp - \frac{1}{2} (x-b)' C^{-1} (x-b).$$

A  $k$ -variate Student  $t$  density on  $x \in \mathbb{R}^k$  with  $r > 0$  degrees of freedom, location vector  $b \in \mathbb{R}^k$  and PDS  $k \times k$  precision matrix  $A$ :

$$f_S^k(x|r,b,A) = \frac{\Gamma(\frac{r+k}{2})}{\Gamma(\frac{r}{2}) (r\pi)^{k/2}} |A|^{\frac{1}{2}} [1 + \frac{1}{r} (x-b)' A (x-b)]^{-\frac{r+k}{2}}.$$

A gamma density on  $z > 0$  with  $a, b > 0$ :

$$f_G(z|a,b) = b^a [\Gamma(a)]^{-1} z^{a-1} \exp(-bz).$$

A beta density on  $v \in (0,1)$  with  $a, b > 0$ :

$$f_B(v|a,b,c) = \frac{\Gamma(a+b)}{c\Gamma(a)\Gamma(b)} \left(\frac{v}{c}\right)^{a-1} \left(1 - \frac{v}{c}\right)^{b-1}.$$

A three-parameter inverted beta or beta prime density on  $w > 0$  with  $a, b, c > 0$  [see Zellner (1971, p. 376)]:

$$f_{IB}(w|a,b,c) = \frac{\Gamma(a+b)}{c\Gamma(a)\Gamma(b)} \left(\frac{w}{c}\right)^{b-1} \left(1 + \frac{w}{c}\right)^{-(a+b)}.$$

## References

- Chib, S., Osiewalski, J., and Steel, M.F.J., 1991, Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models, mimeo.
- Chib, S., Tiwari, R.C., and Jammalamadaka, S.R., 1988, Bayes Prediction in Regressions with Elliptical Errors, *Journal of Econometrics*, 38, 349-360.
- Dickey, J.M., and Chen, C.-H., 1985, Direct Subjective-Probability Modeling Using Ellipsoidal Distributions, in *Bayesian Statistics 2*, eds. J.M. Bernardo, M.H. De Groot, D.V. Lindley, and A.F.M. Smith, Amsterdam: North-Holland, 157-182.
- Florens, J.P., and Mouchart, M., 1985, Conditioning in Dynamic Models, *Journal of Time Series Analysis*, 6, 15-34.
- Jammalamadaka, S.R., Tiwari, R.C., and Chib, S., 1987, Bayes Prediction in the Linear Model with Spherically Symmetric Errors, *Economics Letters*, 24, 39-44.
- Johnson, M.E., 1987, *Multivariate Statistical Simulation*, New York: Wiley.
- Kelker, D., 1970, Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization, *Sankhya, Ser. A*, 32, 419-430.
- Osiewalski, J., 1991, A Note on Bayesian Inference in a Regression Model with Elliptical Errors, *Journal of Econometrics*, 48, 183-193.
- Osiewalski, J., and Steel, M.F.J., 1990, Robust Bayesian Inference in Elliptical Regression Models, *Center Discussion Paper 9032*, Tilburg University.



Richard, J.F., and Steel, M.F.J., 1988, Bayesian Analysis of Systems of Seemingly Unrelated Regression Equations Under a Recursive Extended Natural Conjugate Prior Density, *Journal of Econometrics*, 38, 7-37.

Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley.

Zellner, A., 1976, Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms, *Journal of the American Statistical Association*, 71, 400-405.

**Discussion Paper Series, CentER, Tilburg University, The Netherlands:**

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9005	Th. ten Raa and M.F.J. Steel	A Stochastic Analysis of an Input-Output Model: Comment
9006	M. McAleer and C.R. McKenzie	Keynesian and New Classical Models of Unemployment Revisited
9007	J. Osiewalski and M.F.J. Steel	Semi-Conjugate Prior Densities in Multi- variate t Regression Models
9008	G.W. Imbens	Duration Models with Time-Varying Coefficients
9009	G.W. Imbens	An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling
9010	P. Deschamps	Expectations and Intertemporal Separability in an Empirical Model of Consumption and Investment under Uncertainty
9011	W. Güth and E. van Damme	Gorby Games - A Game Theoretic Analysis of Disarmament Campaigns and the Defense Efficiency-Hypothesis
9012	A. Horsley and A. Wrobel	The Existence of an Equilibrium Density for Marginal Cost Prices, and the Solution to the Shifting-Peak Problem
9013	A. Horsley and A. Wrobel	The Closedness of the Free-Disposal Hull of a Production Set
9014	A. Horsley and A. Wrobel	The Continuity of the Equilibrium Price Density: The Case of Symmetric Joint Costs, and a Solution to the Shifting-Pattern Problem
9015	A. van den Elzen, G. van der Laan and D. Talman	An Adjustment Process for an Exchange Economy with Linear Production Technologies
9016	P. Deschamps	On Fractional Demand Systems and Budget Share Positivity
9017	B.J. Christensen and N.M. Kiefer	The Exact Likelihood Function for an Empirical Job Search Model
9018	M. Verbeek and Th. Nijman	Testing for Selectivity Bias in Panel Data Models
9019	J.R. Magnus and B. Pesaran	Evaluation of Moments of Ratios of Quadratic Forms in Normal Variables and Related Statistics
9020	A. Robson	Status, the Distribution of Wealth, Social and Private Attitudes to Risk

No.	Author(s)	Title
9021	J.R. Magnus and B. Pesaran	Evaluation of Moments of Quadratic Forms in Normal Variables
9022	K. Kamiya and D. Talman	Linear Stationary Point Problems
9023	W. Emons	Good Times, Bad Times, and Vertical Upstream Integration
9024	C. Dang	The $D_2$ -Triangulation for Simplicial Homotopy Algorithms for Computing Solutions of Nonlinear Equations
9025	K. Kamiya and D. Talman	Variable Dimension Simplicial Algorithm for Balanced Games
9026	P. Skott	Efficiency Wages, Mark-Up Pricing and Effective Demand
9027	C. Dang and D. Talman	The $D_1$ -Triangulation in Simplicial Variable Dimension Algorithms for Computing Solutions of Nonlinear Equations
9028	J. Bai, A.J. Jakeman and M. McAleer	Discrimination Between Nested Two- and Three- Parameter Distributions: An Application to Models of Air Pollution
9029	Th. van de Klundert	Crowding out and the Wealth of Nations
9030	Th. van de Klundert and R. Gradus	Optimal Government Debt under Distortionary Taxation
9031	A. Weber	The Credibility of Monetary Target Announce- ments: An Empirical Evaluation
9032	J. Osiewalski and M. Steel	Robust Bayesian Inference in Elliptical Regression Models
9033	C. R. Wichers	The Linear-Algebraic Structure of Least Squares
9034	C. de Vries	On the Relation between GARCH and Stable Processes
9035	M.R. Baye, D.W. Jansen and Q. Li	Aggregation and the "Random Objective" Justification for Disturbances in Complete Demand Systems
9036	J. Driffill	The Term Structure of Interest Rates: Structural Stability and Macroeconomic Policy Changes in the UK
9037	F. van der Ploeg	Budgetary Aspects of Economic and Monetary Integration in Europe
9038	A. Robson	Existence of Nash Equilibrium in Mixed Strategies for Games where Payoffs Need not Be Continuous in Pure Strategies

No.	Author(s)	Title
9039	A. Robson	An "Informationally Robust Equilibrium" for Two-Person Nonzero-Sum Games
9040	M.R. Baye, G. Tian and J. Zhou	The Existence of Pure-Strategy Nash Equilibrium in Games with Payoffs that are not Quasiconcave
9041	M. Burnovsky and I. Zang	"Costless" Indirect Regulation of Monopolies with Substantial Entry Cost
9042	P.J. Deschamps	Joint Tests for Regularity and Autocorrelation in Allocation Systems
9043	S. Chib, J. Osiewalski and M. Steel	Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models
9044	H.A. Keuzenkamp	The Probability Approach in Economic Methodology: On the Relation between Haavelmo's Legacy and the Methodology of Economics
9045	I.M. Bomze and E.E.C. van Damme	A Dynamical Characterization of Evolutionarily Stable States
9046	E. van Damme	On Dominance Solvable Games and Equilibrium Selection Theories
9047	J. Driffill	Changes in Regime and the Term Structure: A Note
9048	A.J.J. Talman	General Equilibrium Programming
9049	H.A. Keuzenkamp and F. van der Ploeg	Saving, Investment, Government Finance and the Current Account: The Dutch Experience
9050	C. Dang and A.J.J. Talman	The D <sub>1</sub> -Triangulation in Simplicial Variable Dimension Algorithms on the Unit Simplex for Computing Fixed Points
9051	M. Baye, D. Kovenock and C. de Vries	The All-Pay Auction with Complete Information
9052	H. Carlsson and E. van Damme	Global Games and Equilibrium Selection
9053	M. Baye and D. Kovenock	How to Sell a Pickup Truck: "Beat-or-Pay" Advertisements as Facilitating Devices
9054	Th. van de Klundert	The Ultimate Consequences of the New Growth Theory; An Introduction to the Views of M. Fitzgerald Scott
9055	P. Kooreman	Nonparametric Bounds on the Regression Coefficients when an Explanatory Variable is Categorized
9056	R. Bartels and D.G. Fiebig	Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads

No.	Author(s)	Title
9057	M.R. Veall and K.F. Zimmermann	Evaluating Pseudo-R <sup>2</sup> 's for Binary Probit Models
9058	R. Bartels and D.G. Fiebig	More on the Grouped Heteroskedasticity Model
9059	F. van der Ploeg	Channels of International Policy Transmission
9060	H. Bester	The Role of Collateral in a Model of Debt Renegotiation
9061	F. van der Ploeg	Macroeconomic Policy Coordination during the Various Phases of Economic and Monetary Integration in Europe
9062	E. Bennett and E. van Damme	Demand Commitment Bargaining: - The Case of Apex Games
9063	S. Chib, J. Osiewalski and M. Steel	Regression Models under Competing Covariance Matrices: A Bayesian Perspective
9064	M. Verbeek and Th. Nijman	Can Cohort Data Be Treated as Genuine Panel Data?
9065	F. van der Ploeg and A. de Zeeuw	International Aspects of Pollution Control
9066	F.C. Drost and Th. E. Nijman	Temporal Aggregation of GARCH Processes
9067	Y. Dai and D. Talman	Linear Stationary Point Problems on Unbounded Polyhedra
9068	Th. Nijman and R. Beetsma	Empirical Tests of a Simple Pricing Model for Sugar Futures
9069	F. van der Ploeg	Short-Sighted Politicians and Erosion of Government Assets
9070	E. van Damme	Fair Division under Asymmetric Information
9071	J. Eichberger, H. Haller and F. Milne	Naive Bayesian Learning in 2 x 2 Matrix Games
9072	G. Alogoskoufis and F. van der Ploeg	Endogenous Growth and Overlapping Generations
9073	K.C. Fung	Strategic Industrial Policy for Cournot and Bertrand Oligopoly: Management-Labor Cooperation as a Possible Solution to the Market Structure Dilemma
9101	A. van Soest	Minimum Wages, Earnings and Employment
9102	A. Barten and M. McAleer	Comparing the Empirical Performance of Alternative Demand Systems



No.	Author(s)	Title
9103	A. Weber	EMS Credibility
9104	G. Alogoskoufis and F. van der Ploeg	Debts, Deficits and Growth in Interdependent Economies
9105	R.M.W.J. Beetsma	Bands and Statistical Properties of EMS Exchange Rates
9106	C.N. Teulings	The Diverging Effects of the Business Cycle on the Expected Duration of Job Search
9107	E. van Damme	Refinements of Nash Equilibrium
9108	E. van Damme	Equilibrium Selection in $2 \times 2$ Games
9109	G. Alogoskoufis and F. van der Ploeg	Money and Growth Revisited
9110	L. Samuelson	Dominated Strategies and Common Knowledge
9111	F. van der Ploeg and Th. van de Klundert	Political Trade-off between Growth and Government Consumption
9112	Th. Nijman, F. Palm and C. Wolff	Premia in Forward Foreign Exchange as Unobserved Components
9113	H. Bester	Bargaining vs. Price Competition in a Market with Quality Uncertainty
9114	R.P. Gilles, G. Owen and R. van den Brink	Games with Permission Structures: The Conjunctive Approach
9115	F. van der Ploeg	Unanticipated Inflation and Government Finance: The Case for an Independent Common Central Bank
9116	N. Rankin	Exchange Rate Risk and Imperfect Capital Mobility in an Optimising Model
9117	E. Bomhoff	Currency Convertibility: When and How? A Contribution to the Bulgarian Debate!
9118	E. Bomhoff	Stability of Velocity in the G-7 Countries: A Kalman Filter Approach
9119	J. Osiewalski and M. Steel	Bayesian Marginal Equivalence of Elliptical Regression Models



PO BOX 00153 5000 LE TILBURG THE NETHERLANDS

**Bibliotheek K. U. Brabant**



17 000 011 17506 5